Markov Chain Monte Carlo (MCMC) Methods

- Markov Chain – stochastic process that has the Markov property

- Monte Carlo ?
  - small town in Monaco with a famous casino (Monte Carlo casino)
  - Monte carlo Methods = simulation of random processes, estimate properties of a distribution by examining random samples from a distn.
    i.i.d.

    e.g. Draw a large # of random samples from a given distribution, calculate the sample mean

def: MCMC methods are a class of algorithms for sampling from a probability distribution based on constructing a MC that has the desired distribution as its ~~equilibrium~~ distribution.

aka stationary

Random samples are not i.i.d.!
correlated since

(Random samples are generated by a sequential process (MC) & uses the Markov property)

each sample only depends on the one before it

Random Walk Monte Carlo methods make up a large subclass of MCMC methods.

Applications

- MCMC methods are primarily used for calculating numerical approximations of multi-dimensional integrals
  i.e. expected values
  e.g. In areas of Bayesian stats, computational physics & biology, others
- Rare event sampling

Examples ← Algorithms for constructing the MC

- Metropolis-Hastings algorithm:
  this method generates a RW using a proposal density & a rule for rejecting some of the proposed moves.
  for new steps
- ~~Gibbs sampler:~~
  this method requires all conditional distributions of the target distribution to be sampled exactly.
  (simulates the posterior distn & enables full Bayesian inference)
  Special case of MH

Markov Chain Monte Carlo (MCMC) Methods (con't)

SKIP

Recall def: Class of algorithms for sampling from a prob. distn. based on constructing a MC that has the desired distn. as its stationary distribution.

2 Main Examples:

① Metropolis-Hastings algorithm
② Gibbs sampler

Metropolis-Hastings MCMC

- Method for obtaining a sequence of random samples from a probability distn for which direct sampling is difficult.
- This sequence can be used to approximate the distn (e.g. generate a histogram) or to compute an integral (e.g. an expected value).
- Generally used for sampling from multi-dimensional distributions (especially when dim. is high).
- Draw samples from any prob. distn. $P(x)$ provided you can compute the value of a function $g(x)$ that is proportional to the density of $P(x)$.

$$g(x) = C\underbrace{f(x)}_{\text{PDF for dist'n } P}$$

1st

Some Examples of MCMC

(1) Suppose we want to sample from a Beta distribution which has PDF:

$$f(x) = \underset{\nearrow}{C} x^{\alpha-1}(1-x)^{\beta-1}$$

maybe ~~unknown~~

, C is the normalizing constant

i.e. C is s.t. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

property of a PDF

* In reality, you might need to work with a more complicated distribution function & often you won't actually know the normalizing constant (OR very hard to compute)

> Here, $C = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, $\Gamma(r)$ – Gamma function

MCMC → create a MC which has the Beta distn as its stationary distn, given that we can sample from a uniform distn (easy!)

Describe Metropolis-Hastings Algorithm & Implement it in R.

2nd

MH algorithm :

Define target (stationary) distribution $\pi(x)$.

Want to create a MC that has this stationary distn.

Start with an arbitrary (irreducible) MC with transition matrix Q

s.t. Q_{ij} is the probability of moving from state i to state j. $Q_{ij} = P(X_{n+1}=j \mid X_n = i)$

1. Start at a random initial state i.

2. Generate a new proposal state by (taking the previous state and adding some random noise) according to a proposal distribution. (Same as [using the i^{th} row of the transition matrix Q.])

   typical choice: Normal centered at current state i → pts closer to i are more likely to be visited next → RW

3. Compute the acceptance probability

$$\alpha_{ij} = \min\left(1, \pi_j Q_{ji} / \pi_i Q_{ij}\right)$$

4. Flip a coin that lands heads with probability α_{ij}.
   (aka Generate uniform random number)

   If coin lands heads up, accept new proposal state.

   If not, stay at the current state.

5. Repeat for a long time!

→ MC will converge to π,
use the states of the chain as the random sample

Metropolis-Hastings (con't)

[More details ... see Simulation §12.2]

- Let $\mathbb{Q}$ be an irreducible Markov transition prob. matrix on $\{1, 2, \dots, m\}$; $\mathbb{Q} = (Q_{ij})$. Target dist'n π.

- Define MC $\{X_n : n \geq 0\}$ as follows:

  when $X_n = i$, generate RV X s.t.

  $P(X=j) = Q_{ij}$ for $j \in \{1, \dots, m\}$.

  - If $X = j$, then set $X_{n+1} = j$ w/prob. α_{ij} OR set $X_{n+1} = j$ w/ prob $1-\alpha_{ij}$

- The sequence of states ~~is~~ will constitute a MC with transition matrix $\mathbb{P}$ s.t.

$$P_{ij} = Q_{ij}\alpha_{ij}, \quad \text{if } j \neq i$$

$$P_{ii} = Q_{ii} + \sum_{k \neq i} Q_{ik}(1-\alpha_{ik})$$

- This MC will have stationary probabilities π_j s.t.

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \text{(aka time reversible)}$$

$$\Longleftrightarrow \quad \pi_i Q_{ij} \alpha_{ij} = \pi_j Q_{ji} \alpha_{ji}$$

This condition is satisfied if we take acceptance probability

$$\alpha_{ij} = \min\left(\frac{\pi_j Q_{ji}}{\pi_i Q_{ij}}, 1\right) = \underbrace{\min\left(\frac{b_j Q_{ji}}{b_i Q_{ij}}, 1\right)}$$

for Motivating example

Do 1st ??

Motivating Example

Let b_j, $j=1,2,\ldots,m$ be positive #'s and let

$B = \sum_{j=1}^{m} b_j$. Suppose m large & B difficult to calculate.

Suppose we want to simulate a sequence of RVs with PMF

$$\pi_j = \frac{b_j}{B}, \quad j=1,2,\ldots,m.$$

<u>One Method</u>: MH algorithm (MCMC method)

Simulate a seq. of RVs whose dist'ns converge to π_j by finding a MC (easy to simulate) whose limiting probabilities are π_j's.

Lecture 13 | Stat 753 | 10/10/17 2

Metropolis-Hastings (con't)

Note: Only time we use the PDF π is to find the acceptance probability α_{ij}

In that, we divide π_j / π_i so the normalizing constant gets cancelled.

e.g. Say, $\pi \sim Beta(\alpha, \beta)$

Beta PDF: $f(x) = C x^{\alpha-1}(1-x)^{\beta-1}$, C cancels when take ratio

$$C = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad \text{where} \quad \Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx$$

Gamma Function

* Works for any $g(x) \propto f(x)$ density function
$g = C \cdot f$

Simple Examples in R

Open mcmc-examples.r in RStudio

Example 1

- Motivated by a professor interested in finding the mean test score in a student population.
- Given that scores are normally distributed with SD=15.
- Use MCMC to draw samples from target distribution: $N(100, 15)$
  (given a single observation: score = 100)

far outside distn

Generate 500 samples

tail of distn

Initial value = 110 (or 250 or 600)

New proposal value = previous state + $\underbrace{\text{random noise}}_{\text{rnorm}(1,0,5)}$

mean SD

uses proposal distribution $Q_{ij} = P(X_{n+1}=j \mid X_n=i)$

Compare height of target distribution at proposal value
to " " " current value.

i.e. $$\frac{\text{dnorm}(\text{proposal}, \mu=100, sd=15)\cdot \cancel{Q_{ji}}}{\text{dnorm}(\underbrace{\text{samples[i-1]}}_{\text{current state}}, \mu=100, sd=15)\cdot \cancel{Q_{ij}}}$$

Generate random uniform(0,1) number, u.

If ratio is bigger than u, accept new proposed state.

(Here $Q_{ij} = Q_{ji}$ so they cancel)
since proposal distn is symmetric (Normal dist'n)

Example 2

Similar, but target distn = $N(0,1)$ Standard Normal
& proposal distn = $U(-\alpha,\alpha)$ Uniform